Suitability of Non-Parametric Stereo Matching Techniques for Mining Automation

Jasmine Banks[1,2], Mohammed Bennamoun*[1], Kurt Kubik[1] and Peter Corke[2,3]

*Corresponding author
1: Space Centre for Satellite Navigation, Queensland University of Technology
2: Centre for Mining Technology and Equipment
3: CSIRO Manufacturing Science and Technology

The mining environment, being complex, irregular and time varying, presents a challenging prospect for stereo vision. For this application, speed, reliability, and the ability to produce a dense depth map are of foremost importance. This paper assesses the suitability of a number of matching techniques for use in a stereo vision sensor for close range scenes consisting primarily of rocks. These include traditional area-based matching metrics, and non-parametric transforms, in particular, the rank and census transforms. Experimental results show that the rank and census transforms exhibit a number of clear advantages over area-based matching metrics, including their low computational complexity, and robustness to certain types of distortion.

Suitability of Non-Parametric Stereo Matching Algorithms for Mining Automation

Jasmine Banks[1,2], Mohammed Bennamoun[1], Kurt Kubik[1] and Peter Corke[2,3]

[1] Space Centre for Satellite Navigation,Queensland University of Technology
Brisbane 4000, Australia.
[2] Cooperative Research Centre for Mining Technology and Equipment
[3] CSIRO Manufacturing Science and Technology
{j.banks, m.bennamoun, k.kubik}@qut.edu.au, pic@cat.csiro.au

Abstract

*The mining environment, being complex, irregular and time varying, presents a challenging prospect for **stereo vision**. For this application, speed, reliability, and the ability to produce a dense depth map are of foremost importance. This paper assesses the suitability of a number of **matching** techniques for use in a stereo vision sensor for close range scenes consisting primarily of rocks. These include traditional **area-based** matching metrics, and **non-parametric** transforms, in particular, the **rank** and **census** transforms. Experimental results show that the rank and census transforms exhibit a number of clear advantages over area-based matching metrics, including their low computational complexity, and robustness to certain types of distortion.*

1. Introduction

Perception of the three-dimensional environment is a pre-requisite for mine equipment automation, since autonomous vehicles and robot devices need to be aware of the surrounding environment in order to plan their actions and carry out tasks. Stereo vision is a technique used to discern depth information from a scene, in which two (or more) images of a scene are taken from different perspectives, and depth is computed from stereo disparity. Applications for stereo vision include aerial photogrammetry[19,23], autonomous vehicle guidance [21], robotics and industrial automation. The mining environment, being complex, irregular and time varying, presents a challenging prospect for stereo vision. For this application, speed, reliability, and the ability to produce a dense depth map are of foremost importance[11].

A fundamental issue is to establish *correspondence* or *matching* of points in two images, such as the ROCK stereo pair of Figure 1, in order to compute the disparity and subsequently the depth information. This paper assesses the suitability of a number of matching techniques for use in a stereo vision sensor for close range scenes consisting primarily of rocks.

Matching techniques may typically be categorised according to the type of matching primitives they employ[14]:

Area-based which are distinguished by the fact that actual grey-level pixel values in the stereo images are compared to find the best match. The information contained in a single pixel is not sufficient for unambiguous matching, therefore regularly sized pixel neighbourhoods are compared using matching metrics[13,14,18].

Transform-based in which some manner of transformation of the pixel values in the stereo images takes place prior to matching. The transformed images may then be matched using area-based metrics. Examples include filtering[4], and non-parametric transforms including the rank and census transforms[25].

Feature-based which are characterised by the use of image features such as edges, vertices and polygons as the matching primitives. These methods rely on feature extraction[7]. The symbolic representations of these features are then compared to find the best match[2,16,22].

Feature-based matchers tend to be faster than area-based methods, since only a small subset of pixels are used. However, they typically yield very sparse depth maps, since matching only takes place at image locations where features occur, and results for intermediate points must be obtained by interpolation[16,20]. This interpolation process relies on assumptions about the scene geometry between features. Feature-based matchers are also highly accurate since features may be located with sub-pixel precision. They are best suited to images where features are relatively sparse, such as scenes containing planar surfaces delineated by edges. Such scenes would typically be comprised of man-made objects. Area-based matchers are usually unsuitable to use on these images, since their smooth surfaces lack sufficient texture for an area-based matcher to match on.

Area based techniques, on the other hand, are best suited to highly textured scenes, in contrast to feature-based techniques which tend to be confused by a large amount of surface texture[18]. Area-based matchers can also potentially yield matching results for every image pixel

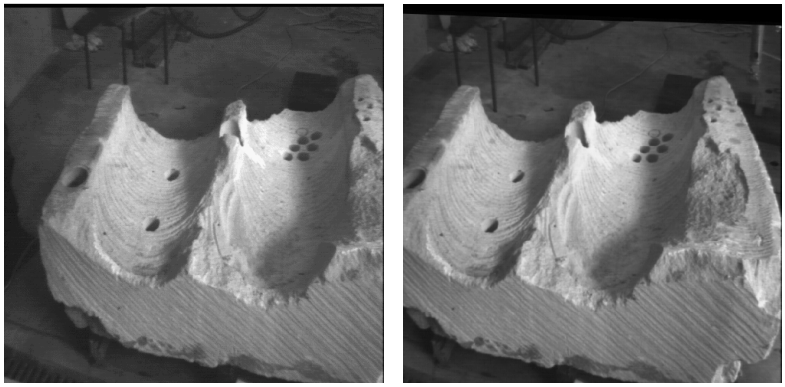


Figure 1: ROCK stereo pair.

and hence yield a dense depth map. The advantages of area-based algorithms include their simplicity and straightforward implementation, as well as their amenability to hardware realisation[1]. However, their accuracy is not as high as the feature-based methods. This is due to the smoothing effect introduced by using a square window of pixels for matching[10].

This paper examines area-based matching techniques in detail, for the following reasons:

1. Scenes comprised of rocks usually have a large amount of surface texture, and are therefore well suited to area based matching.
2. They have the potential to yield a dense depth map.
3. They are amenable to fast hardware implementation.

Transform-based techniques, being based on the matching of dense information, essentially exhibit the same advantages and disadvantages as area-based techniques. A class of transforms known as non-parametric transforms have recently been proposed for stereo matching. In addition to the advantages listed above, these transforms are robust to radiometric distortion and small amounts of random noise. The objective of this paper is to assess their suitability for mining automation applications.

2. Area-Based Matching

In area-based matching, a point to be matched essentially becomes the centre of a small window of pixels, and this window is compared with similarly sized regions in the other image. *Matching metrics* are used to provide a numerical measure of the similarity between a template window in the first image and a candidate window in the second image, and hence are used to determine the optimum match.

Epipolar geometry[6] is used to improve the efficiency of the matching process by constraining the search to one dimension. Stereo images may be rectified such that the epipolar lines correspond to the horizontal scan lines[3]. A simple approach used in area based matching is to compute the value of the matching metric using a fixed window in the first image and a shifting window in the second image, as illustrated in Figure 2.

The shifting window is moved in integer increments along the epipolar line, where the amount of shift is the test disparity. The disparity having the optimum value for the matching metric is then selected.

2.1 Matching Problems

All area-based matching algorithms must deal with at least the following problems:

Occlusions caused by portions of a scene being visible in only one image.

Repetitive patterns which can potentially result in invalid matches.

Bland regions which do not contain enough information for matching, eg, a featureless wall.

Perspective distortion which occurs because the shape of objects will change when they are viewed from different vantage points.

Radiometric distortion which may result in a constant offset between pixel values in the two images, and/or pixel intensities in one image being multiplied by a gain factor with respect to the other image. These effects are caused by differences in camera parameters, such as gain, bias and gamma factor.

Specular reflection caused by the reflectance properties of the object. Matching algorithms usually assume Lambertian reflection model, in which an object reflects light equally in all directions. It is therefore assumed that a particular point will have the same intensity regardless of the direction from which it is viewed. However, this is often not the case, with specular (mirror-like) reflection being the most dramatic departure from the Lambertian case.

Noise which is introduced by the image acquisition and digitisation process.

2.2 Matching Metrics

A number of classical matching metrics are listed in Table I. All these metrics use a square window of pixels as the basis for comparison.

SAD	Sum of Absolute Differences	$\sum_{(u,v)\in W} \left| I_1(u,v) - I_2(u,v) \right|$
ZSAD	Zero Mean Sum of Absolute Differences	$\sum_{(u,v)\in W} \left| \left(I_1(u,v) - \overline{I_1(u,v)}\right) - \left(I_2(u,v) - \overline{I_2(u,v)}\right) \right|$
SSD	Sum of Squared Differences	$\sum_{(u,v)\in W} \left(I_1(u,v) - I_2(u,v)\right)^2$
ZSSD	Zero Mean Sum of Squared Differences	$\sum_{(u,v)\in W} \left(\left(I_1(u,v) - \overline{I_1(u,v)}\right) - \left(I_2(u,v) - \overline{I_2(u,v)}\right)\right)^2$
NCC	Normalised Cross Correlation	$\frac{\sum_{(u,v)\in W} I_1(u,v) \cdot I_2(u,v)}{\sqrt{\sum_{(u,v)\in W} {I_1}^2(u,v) \cdot \sum_{(u,v)\in W} {I_2}^2(u,v)}}$
ZNCC	Zero Mean Normalised Cross Correlation	$\frac{\sum_{(u,v)\in W} \left(I_1(u,v) - \overline{I_1(u,v)}\right) \cdot \left(I_2(u,v) - \overline{I_2(u,v)}\right)}{\sqrt{\sum_{(u,v)\in W} \left(I_1(u,v) - \overline{I_1(u,v)}\right)^2 \cdot \sum_{(u,v)\in W} \left(I_2(u,v) - \overline{I_2(u,v)}\right)^2}}$

Table I: Area-based matching measures[1]. In each case, I_1 denotes the template window, I_2 denotes the candidate window, and $\sum_{(u,v)\in W}$ denotes summation over the window.

The SAD and the SSD are intuitively the simplest, and computationally the least expensive of all the matching measures[17]. Two areas which consist of exactly the same pixel values would yield a score of zero. However, these measures will no longer yield the correct results in the case of radiometric distortion. The ZSAD and the ZSSD have been devised to deal with this problem, by subtracting the mean of the match area from each intensity value. However, the improved performance of the ZSAD and ZSSD over the SAD and SSD is offset by substantially increased computational complexity.

The NCC measure deals with a possible gain factor by dividing by the variances of each window, while the ZNCC measure additionally deals with the offset problem by first subtracting the mean from each pixel value. For grey level images, these metrics will have a value ranging from -1 to 1, where 1 represents the best match.

2.3 Validation of Matches

Once the optimum match is selected using a matching metric, a number of simple validation techniques may be applied in order to identify incorrect matches.

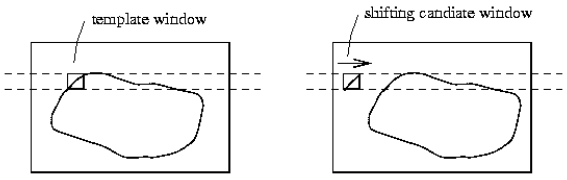


Figure 2: Epipolar constrained area-based matching.

One such technique is left-right consistency checking[14,18], which involves reversing the roles of the two images and performing matching a second time, as illustrated by Figure 3.

Firstly, epipolar constrained matching is carried out using a template window centred on I_1, and the point I_2, which is the best match for I_1, is found. Matching is then performed again, this time using a template window centred on I_2. If this match leads back to the original point I_1, then the match is consistent, otherwise, it is flagged as inconsistent. This validity test is likely to detect invalid matches which may result from bland areas, and also from occlusions. The pixels which comprise an occluded area are likely to match, more or less at random, with locations in the other image. However, these locations are unlikely to match back to the pixels in the occlusion area, rather, they are more likely to match with their own corresponding points. This validation technique can be fooled by repetitive patterns.

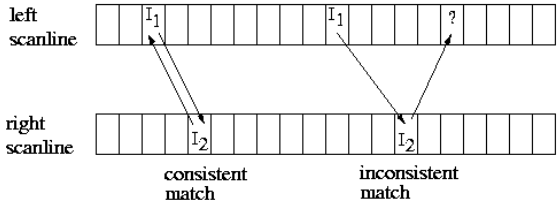


Figure 3 : Consistent and inconsistent matches. The match on the left is consistent, while the match on the right is inconsistent[14].

The number of correct matches can be further increased by removing isolated matches from the matches which remain after left-right consistency checking. This heuristic is based on the assumption that isolated matches are more likely to be incorrect[13,14].

3. Non-Parametric Techniques

Non-parametric techniques are based on the relative ordering of pixel intensities within a window, rather than the intensity values themselves. Consequently, these techniques are robust with respect to radiometric distortion, since differences in gain and bias between two images will not affect the ordering of pixels within a window. In addition, these transforms are tolerant to a small number of outliers within a window, and are therefore robust with respect to small amounts of random noise[8].

Two non-parametric transforms which are suited to fast implementation are[25]:

Rank Transform

This is defined as the number of pixels in the window whose value is less than the centre pixel. The images will therefore be transformed into an array of integers, whose value ranges from 0 to *N-1*, where *N* is the number of pixels in the window. A pair of rank transformed images are then matched using one of the matching metrics of Table I. For hardware implementation, it is advantageous to use a matching metric based on integer arithmetic, such as the SAD or the SSD.

Census Transform

This transform maps the window surrounding the centre pixel to a bit string. If a particular pixel's value is less than the centre pixel then the corresponding position in the bit string will be set to 1, otherwise it is set to zero. Two census transformed images are compared using a similarity metric based on the Hamming distance, ie, the number of bits that differ in the two bit strings. The Hamming distance is summed over the window, ie,

$$\sum_{(u,v)\in W} \text{Hamming}(I_1'(u,v), I_2'(x+u, y+v)) \quad \textbf{(1)}$$

where I_1' and I_2' represent the census transforms of I_1 and I_2. Two hardware implementations of this scheme are discussed in [12,24].

4. Experimental Results

A matching scheme used to compare various area-based metrics is shown in Figure 4.

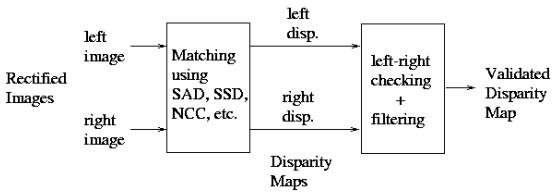


Figure 4: Overall matching process using area-based matching metrics.

In each case, a rectified stereo pair is input to the matching stage, which uses one of the metrics from Table I to determine the initial disparity maps with respect to each image. The left-right consistency criterion, in addition to filtering to remove isolated matches, are then applied, in order to remove invalid matches. The resulting output consists of a disparity map with respect to the right image, from which invalid matches have been removed.

The steps involved in matching using the rank and census transforms are shown in Figure 5 and Figure 6. respectively. The rank transformed stereo images are matched using the SAD metric, while the census transformed images are matched using the Hamming measure of Equation (1). In each case, the disparity maps output from the matcher may then be input to the validity checking stage of Figure 4.

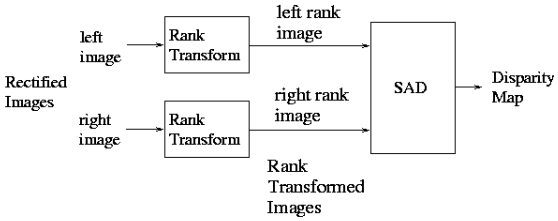


Figure 5: Overall matching process rank transform.

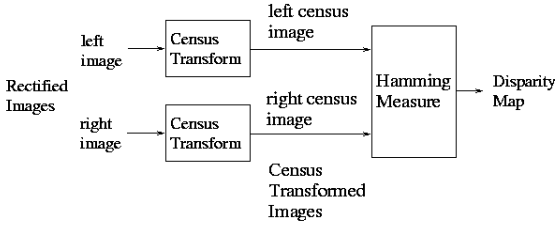


Figure 6: Overall matching process census transform.

The algorithms of Figure 4, Figure 5 and Figure 6 were tested using a large number of test stereo pairs, including the ROCK, IROCKS1 and J1 pairs of Figure 1, Figure 9 and Figure 12 respectively. The IROCKS1 and J1 test pairs were used in the JISCT stereo evaluation[9], and are both affected by radiometric distortion. In IROCKS1, the left image is approximately 28% brighter than the right, while in J1, the right image is approximately 13% brighter than the left.

The disparity maps obtained for the ROCK, IROCKS1 and J1 stereo pairs, using the area-based metrics of Table I are

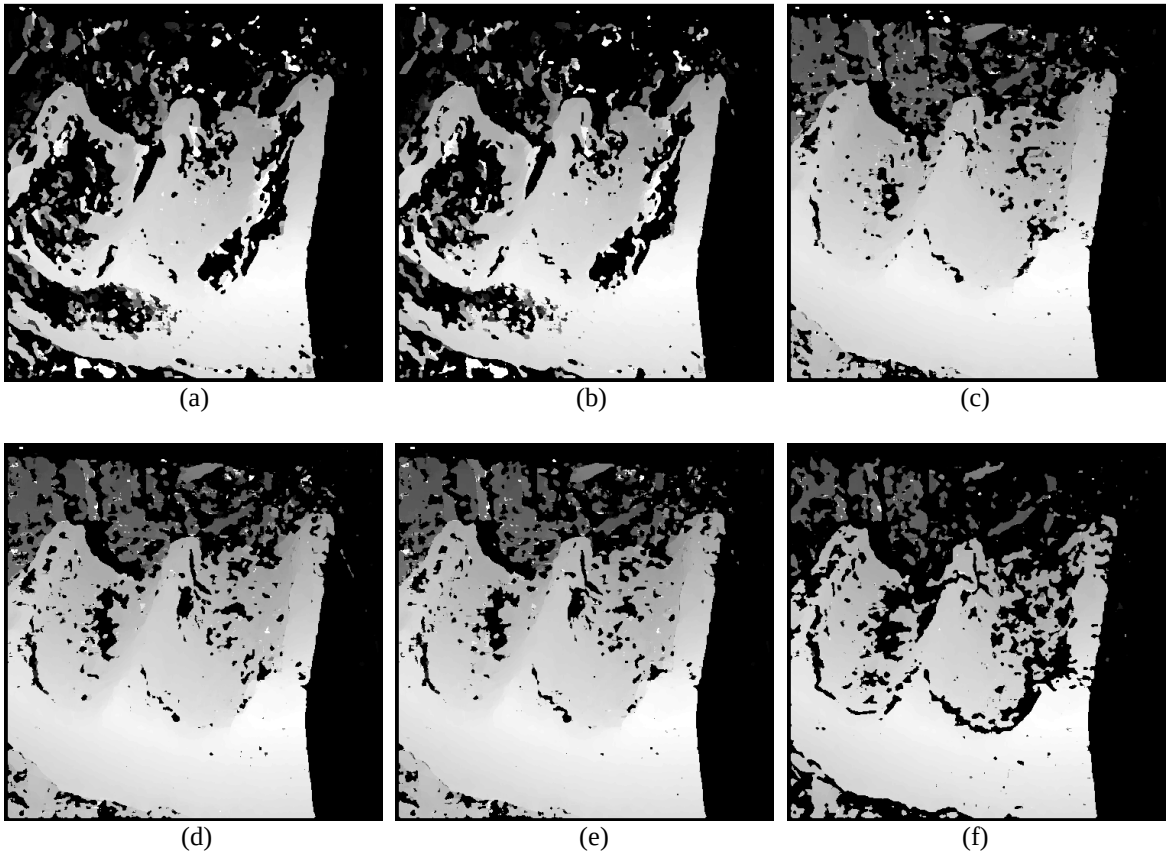


Figure 7: Disparity of ROCK stereo pair, produced using (a) SAD, (b) SSD, (c) NCC, (d) ZSAD, (e) ZSSD, (f) ZNCC metric. The ZSAD, ZSSD, NCC and ZNCC metrics result in the highest proportion of valid matches, however, these metrics have a significantly higher computational overhead than the SAD and the SSD.

shown in Figure 7, Figure 10 and Figure 13 respectively. Lighter regions in the result disparity maps correspond to larger disparities, while black regions consist of invalid matches which were removed. A matching window size of 11 x 11 was used for each metric. The disparity results using the rank and census transforms on the test pairs are shown in Figure 8, Figure 11 and Figure 14 respectively. The census transform was performed using windows of size 5 x 5, however, the matching process used windows of size 11 x 11.

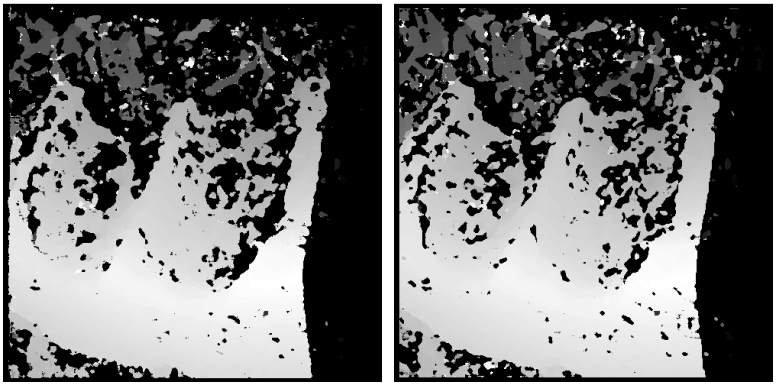
The proportion of matches remaining after validity checking for each metric are shown in Table II. These values represent a preliminary estimate of the performance of each matching metric.

5. Discussion

It can be seen from Figure 7, Figure 10 and Figure 13 that the SAD and the SSD are clearly not robust with respect to radiometric distortion. Use of ZSAD, ZSSD, NCC and ZNCC resulted in improved robustness to radiometric distortion and consequently a higher proportion of valid matches, as shown in Table II. However, these metrics result in increased computational complexity, since they consist of floating point operations. The NCC and ZNCC are particularly computationally expensive due to the presence of floating point multiplication, division and square root operations.

The proportion of matched pixels as shown in Table II is highly dependent on the content of the images. For example, stereo pairs containing large occluded regions would lead to a lower proportion of matched pixels for this pair. Also, the presence of large bland regions, for example, a background wall, can further decrease the proportion of matched pixels. Despite these perturbations, results for all test stereo pairs show that the SAD and the SSD are consistently out-performed by all the other matching metrics tested, as well as the rank and census transforms.

Two matching algorithms based on non-parametric transforms have been tested – the rank transform followed by matching with the SAD metric, and the census transform followed by matching with the Hamming metric. For the test imagery used, matching using the rank and census transforms was found to be reliable in the presence of radiometric distortion. The results of Figures 11 and 14 clearly illustrate the improvement in the disparity results of the rank and census transforms in comparison to the SAD ans SSD results of Figures 10 and 13. Table II shows that, for the test pairs used in this paper, the rank and census algorithms resulted in disparity maps whose proportion of valid matches is comparable to the ZSAD, ZSSD and NCC.



(a) (b)

Figure 8: Disparity of rock stereo pair, produced using (a) Rank transform followed by SAD and (b) Census transform followed by Hamming metric. The rank and census methods result in a higher proportion of valid matches than the SAD and SSD, and in addition, they do not introduce the computational overhead of the ZSAD, ZSSD and ZNCC.

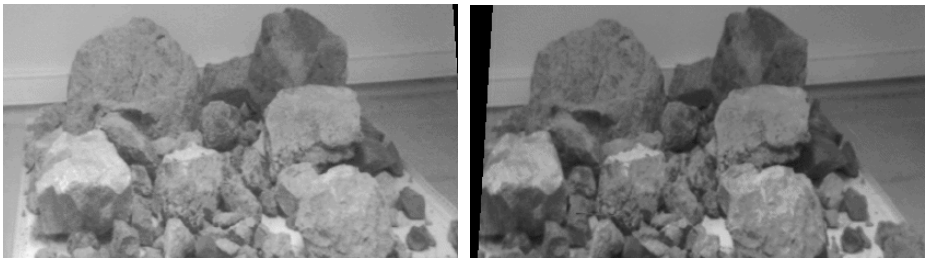


Figure 9: IROCKS1 stereo pair. Note the radiometric distortion, the left image being approximately 28% brighter than the right.

6. Conclusion

This paper has explored the suitability of matching algorithms for a stereo vision sensor for mining automation. The requirements of this sensor are speed, reliability and the ability to produce a dense depth map. Area-based matching techniques have been investigated for this application, for a number of reasons. Firstly, they are suited to textured scenes, and scenes of rocks tend to be textured. Secondly, they have the potential to yield a dense depth map. Finally, they are amenable to real-time hardware implementation.

Previous studies[1] have compared traditional area based metrics for a range of image types. This paper differs from previous work in that in addition to area based metrics, two non-parametric transforms, namely the rank and census, have been investigated. Also, since the purpose of this study was to assess matching algorithms for mining automation, test stereo pairs consisting of close range scenes of rocks were used.

Both the rank and the census transforms were found to result in improved reliability of matching in the presence of radiometric distortion. This is significant since radiometric distortion is a problem which often arises in practice, particularly when low cost cameras are used. In fact, the performance of matching using these transforms is comparable to that of area based metrics such as the ZSAD, ZSSD, and NCC. However, the rank and census transforms have the additional advantage in that they do not introduce the computational complexity of these metrics. Both transforms are also amenable to fast hardware implementation[5], making them potentially suitable for real-time applications. As a result, they merit further investigation for a real-time, reliable stereo matching sensor for mining automation applications.

7. References

[1] Aschwanden, P. and Guggenbühl, W. “Experimental results from a comparative study on correlation-type registration algorithms” In Robust Computer Vision (Förstner, Ruwiedel, Eds.), Wickmann, 1993, pp. 268-289.

[2] Ayache, N. and Faverjon, B. “Efficient registration of stereo images by matching graph descriptions of edge segments”, International Journal of Computer Vision, pp. 107-131, 1987.

[3] Ayache, N. “Artificial Vision for Mobile Robots”, MIT Press, 1991.

[4] Ballard, D. and Brown, C., “Computer Vision”, Prentice Hall, 1982.

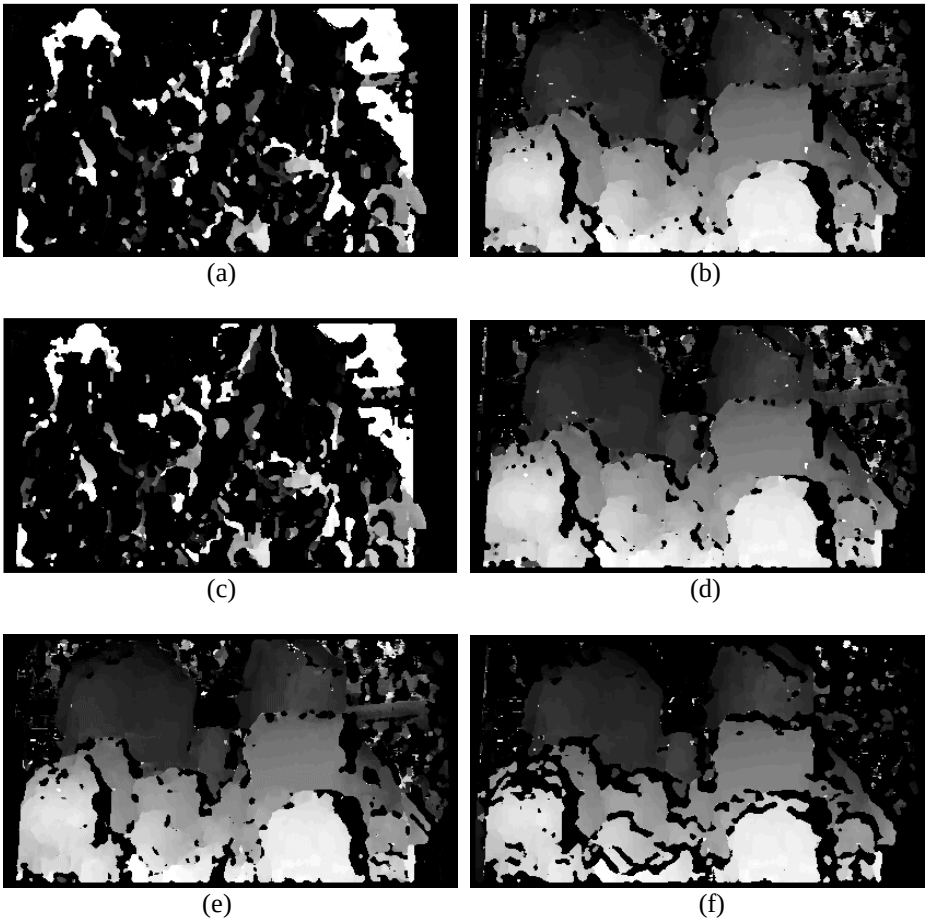


Figure 10: Disparity of IROCKS1 stereo pair, produced using the (a) SAD, (b) SSD, (c) NCC, (d) ZSAD, (e) ZSSD, and (f) ZNCC metrics. Note the poor performance of the SAD and the SSD, due to radiometric distortion. While the ZSAD, ZSSD, NCC and ZNCC result in improved robustness, they introduce additional computational complexity.

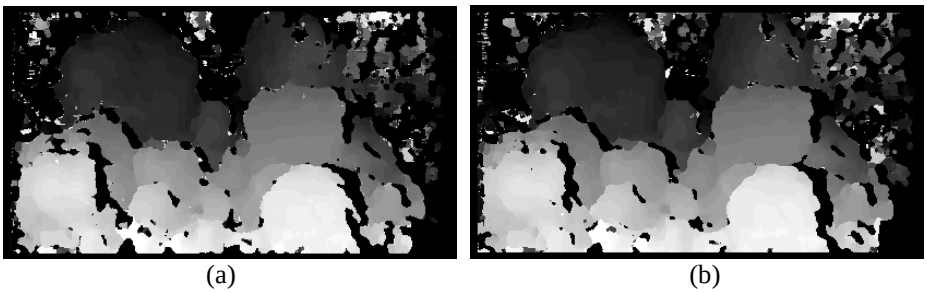


Figure 11: Disparity of IROCKS1 stereo pair, produced using (a) Rank transform followed by SAD and (b) Census transform followed by the Hamming metric. The improvement of the rank and census results over the SAD and SSD clearly illustrates the robustness of the rank and census transforms to radiometric distortion. An additional advantage of the rank and census methods is that they do not introduce the computational overhead of the ZSAD, ZSSD, NCC and ZNCC.

[5] Banks, J., Porter, R., Bennamoun, M. and Corke, P. "A generic implementation framework for stereo matching algorithms", In Proceedings DICTA'97, Auckland, NZ, pp. 29-34, Dec 1997.

[6] Barnard, S. and Fischler, M. "Stereo vision", In Encyclopedia of Artificial Intelligence (S. Shapiro, Ed.), John Wiley & Sons, 1987, pp. 1083-1090.

[7] Bennamoun, M. "Edge detection: problems and solutions", Invited paper in Proceedings of the IEEE Conference on Systems, Man and Cybernetics, Orlando, Florida, Oct 1997.

[8] Bhat, D. and Nayar, S. "Ordinal measures for image correspondence", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 4, pp. 415-423, Apr 1998.

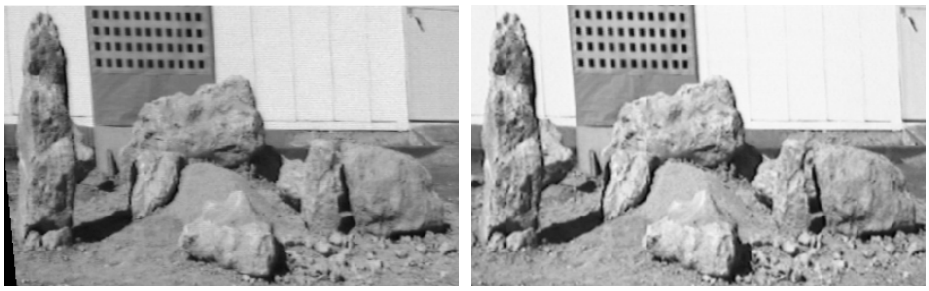
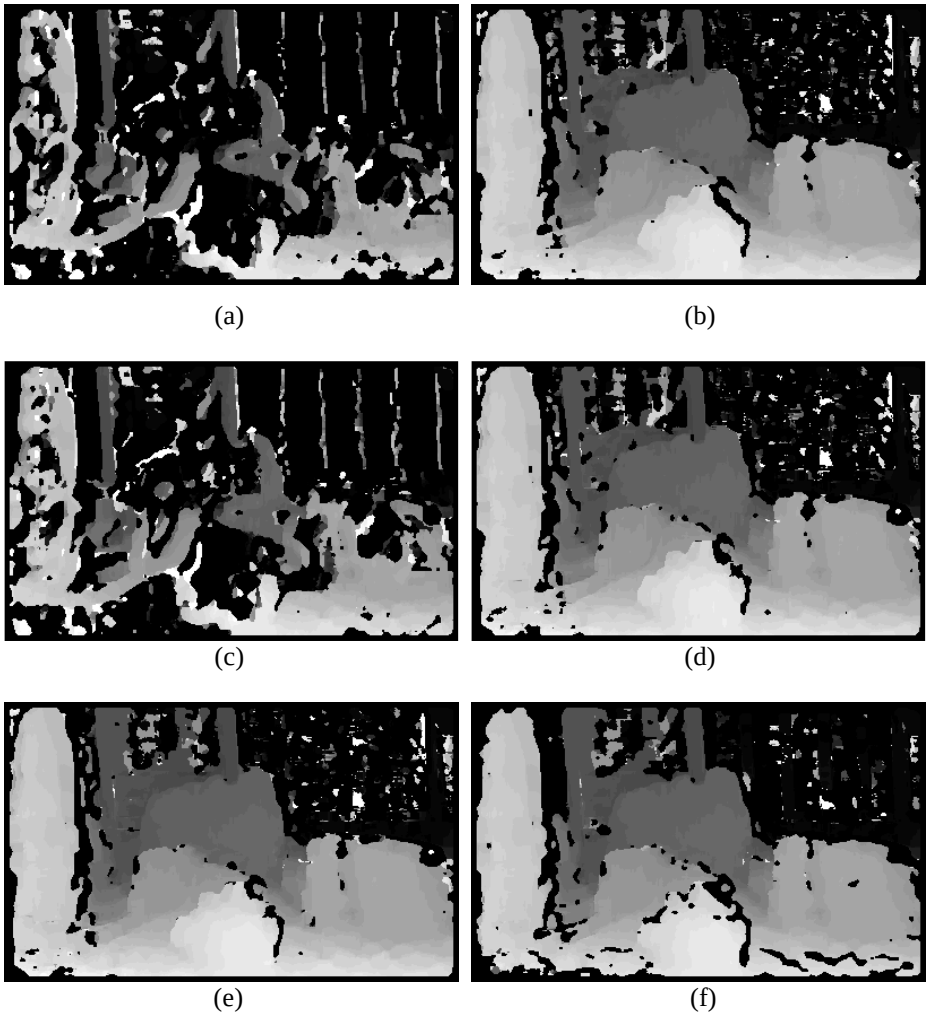


Figure 12: J1 stereo pair. The right image is approximately 13% brighter than the left.



(a) (b)

(c) (d)

(e) (f)

Figure 13: Disparity of J1 stereo pair, produced using (a) SAD, (b) ZSAD, (c) SSD, (d) ZSSD, (e) NCC and (f) ZNCC metrics. As with Figure 11, the poor performance of the SAD and SSD is due to radiometric distortion.

[9] Bolles, R., Baker, H. and Hannah, M. "The JISCT stereo evaluation", Image Understanding Workshop, DARPA, pp. 263-274, 1993.

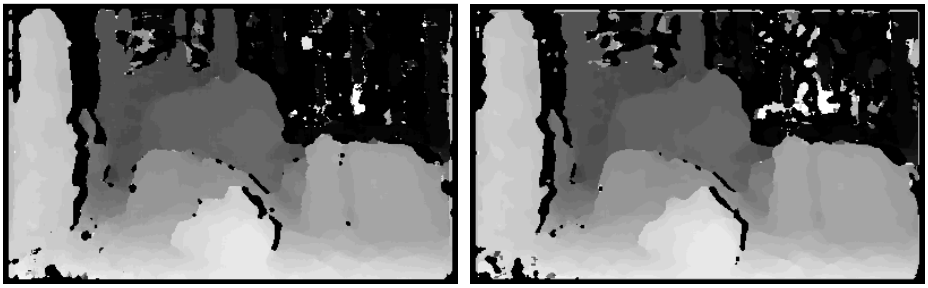
[10] Cochran, S. and Medioni, G. "3-D Surface description from binocular stereo", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, No. 10, pp. 981-994, Oct 1992.

[11] Corke, P. and Winstanley, G. and Roberts, J. "Sensors and control for mining robotics", In Proc. Fourth International Symposium on Mine Mechanisation and Automation, Brisbane, Australia, pp. B1-11-B1-21, Jul 1997.

[12] Dunn, P. and Corke, P. "Real-time stereopsis using FPGAs", In FPGA'97, Imperial College London, Sep 1997.

[13] Faugeras, O., et al. "Real-Time Correlation-Based Stereo: Algorithm, Implementations and Applications", Technical Report 2013, INRIA, 1993.

[14] Fua, P. "A parallel stereo algorithm that produces dense depth maps and preserves image features", Machine Vision and Applications, Vol. 6, pp. 35-49, 1993.



(a) (b)

Figure 14: Disparity of J1 stereo pair, produced using (a) Rank transform followed by SAD and (b) Census transform followed by Hamming metric. As with Figure 10, the rank and census transforms result in improved robustness in the case of radiometric distortion, without introducing the computational complexity of the ZSAD, ZSSD, NCC and ZNCC.

	SAD	ZSAD	SSD	ZSSD	NCC	ZNCC	RANK+SAD	CENSUS
ROCK	0.48	0.65	0.49	0.65	0.65	0.65	0.52	0.61
IROCKS1	0.22	0.71	0.22	0.71	0.69	0.55	0.71	0.77
J1	0.40	0.75	0.46	0.75	0.76	0.72	0.77	0.81

Table II : Proportion of matched pixels for each matching metric.

[15] Greenfeld, J. and Schenk, A. "Experiments with edge-based stereo matching", Photogrammetric Engineering and Remote Sensing, Vol. 55, No. 12, pp. 1771-1777, Dec 1989.

[16] Grimson, W. "From Images to Surfaces", MIT Press, 1981.

[17] Hannah, M. "Computer Matching of Areas in Stereo Images", PhD Thesis, Stanford University, 1974.

[18] Hannah, M. "A system for digital stereo image matching", Photogrammetric Engineering and Remote Sensing, Vol. 55, No. 12, pp. 1765-1770, Dec 1989.

[19] Hsieh, Y., McKeown Jr., D. and Perlant, F., "Performance evaluation of scene registration and stereo matching for cartographic feature extraction", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, No. 2, pp. 214-238, Feb 1992.

[20] Li, S., MRF "Modeling in Computer Vision", Springer-Verlag, 1995.

[21] Matthies, L. "Stereo vision for planetary rovers: stochastic modeling to near real-time implementation", International Journal of Computer Vision, Vol. 8, No. 1, pp. 71-91, 1992.

[22] Medioni, G. and Nevatia, R. "Segment-based stereo matching", Computer Vision, Graphics and Image Processing, Vol. 31, pp. 2-18, 1985.

[23] Wang, Z., "Principles of Photogrammetry (with Remote Sensing)", Wuhan Technical University of Surveying and Mapping, 1990.

[24] Woodfill, J. and Herzen, B. "Real-time stereo vision on the PARTS reconfigurable computer", In Proc. IEEE Symposium in FPGAs for Custom Computing Machines, Los Alamitos, CA, pp. 201-210, Apr 1997.

[25] Zabih, R. and Woodfill, J. "Non-parametric local transforms for computing visual correspondence", In Proc. 3rd European Conference on Computer Vision, Stockholm, 1994.